


[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

[Advanced Scholar Search](#)
[Scholar Preferences](#)
[Scholar Help](#)

Scholar [All articles](#) - [Recent articles](#) Results 1 - 10 of about 58,200 for [inverse document frequency](#)

All Results

[G Salton](#)
[C Buckley](#)
[K Church](#)
[T Joachims](#)
[M Persin](#)

[Inverse document frequency \(IDF\): A measure of deviations from Poisson - all 14 versions »](#)

KW Church, WA Gale - Proceedings of the Third Workshop on Very Large Corpora, 1995 - acl.ldc.upenn.edu

Page 1. **Inverse Document Frequency (IDF): A Measure of Deviations from Poisson ...** We will use **inverse document frequency (IDF)**, a quantity borrowed from ...

Cited by 31 - [Related Articles](#) - [View as HTML](#) - [Web Search](#)

[Text categorization based on weighted inverse document frequency - all 2 versions »](#)

T Tokunaga, M Iwayama - SIG-IPS Japan, 1994 - citeseer.ist.psu.edu

This paper proposes a new term weighting method called weighted **inverse document frequency** WIDF As its name indicates, WIDF is an extension of IDF **inverse ...**

Cited by 20 - [Related Articles](#) - [Cached](#) - [Web Search](#)

[A vector space model for automatic indexing - all 13 versions »](#)

G Salton, A Wong, CS Yang - Communications of the ACM, 1975 - portal.acm.org

... [2] Specifically, if df_k is the **document frequency** of term k , the **inverse document frequency** IDF_k of term k may be defined as [3]: $(IDF)_k = \log_2 n - \log_2 df_k$...

Cited by 788 - [Related Articles](#) - [Web Search](#)

[A comparison of search term weighting: term relevance vs. inverse document frequency - all 2 versions »](#)

H Wu, G Salton - Proceedings of the 4th annual international ACM SIGIR ..., 1981 - portal.acm.org

... A comparison of search term weighting: term relevance vs. **inverse document frequency**. Full text, pdf formatPdf (563 KB). Source, ACM ...

Cited by 13 - [Related Articles](#) - [Web Search](#)

[Understanding inverse document frequency: on theoretical arguments for IDF - all 6 versions »](#)

S Robertson - Journal of Documentation, 2004 - emeraldinsight.com

... The measure of term specificity first proposed in that paper later became known as **inverse document frequency**, or IDF; it is based on counting the number of ...

Cited by 15 - [Related Articles](#) - [Web Search](#) - [BL Direct](#)

[\[book\] A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization - all 26 versions »](#)

T Joachims - 1996 - ldv.uni-trier.de

... once. The **inverse document frequency** $IDF(w_i)$ can be calculated from the **document frequency**. $IDF(w_i) = \log \frac{D}{df(w_i)}$...

Cited by 486 - [Related Articles](#) - [View as HTML](#) - [Web Search](#) - [Library Search](#) - [BL Direct](#)

[Content-based image retrieval with relevance feedback in MARS - all 8 versions »](#)

Y Rui, TS Huang, S Mehrotra - ... Processing, 1997. Proceedings., International Conference on, 1997 - ieeexplore.ieee.org

... This introduces the well-known **inverse document frequency** (idf), which varies inversely with the number of documents in which a term appears. ...

[Cited by 354](#) - [Related Articles](#) - [Web Search](#)

A comparison of Two-Poisson, inverse document frequency and discrimination value models of document ... - all 3 versions »

P Srinivasan - Information Processing and Management: an International ..., 1990 - portal.acm.org

... A comparison of Two-Poisson, **inverse document frequency** and discrimination value models of **document** representation. Source, Information ...

Cited by 10 - Related Articles - Web Search

[Digital Libraries and Autonomous Citation Indexing - all 27 versions »](#)

NEC Steve Lawrence, NEC Kurt Bollacker - Contact, 1999 - doi.ieeeecs.org

... string. Word **frequency** measurements such as term **frequency** \times inverse document **frequency** (TFIDF) are common in information retrieval. ...

Cited by 408 - Related Articles - Web Search - BL Direct

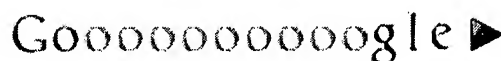
Using Suffix Arrays to Compute Term Frequency and Document Frequency for All Substrings in a Corpus - all 26 versions »

M Yamamoto, KW Church - Computational Linguistics, 2001 - acl.ldc.upenn.edu

... Residual **inverse document frequency (RIDF)** compares document **frequency** to another

model of chance where terms with a particular term **frequency** are ...

[Cited by 62](#) - [Related Articles](#) - [View as HTML](#) - [Web Search](#) - [BL Direct](#)



Result Page: 1 2 3 4 5 6 7 8 9 10 **Next**

inverse document frequency

[Google Home](#) - [About Google](#) - [About Google Scholar](#)

©2007 Google

[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [Gmail](#) [more ▾](#)[Sign in](#)

Google

N-gram-based text categorization

Search

[Advanced Search](#)
[Preferences](#)

Web

Results 1 - 10 of about 19,900 for **N-gram-based text categorization**. (0.13 seconds)**N-Gram-Based Text Categorization - William, John, Trenkle ...**

Text categorization is a fundamental task in document processing, allowing the automated handling of enormous streams of documents in electronic form.

citeseer.ist.psu.edu/68861.html - 22k - [Cached](#) - [Similar pages](#)

[PDF] N-gram based text categorization

File Format: PDF/Adobe Acrobat - [View as HTML](#)

The inspiration for this work comes from the article **N-gram Based Text Categorization** by J. Trenkle and W. Cavnar [1]. They introduced a very simple method ...

www.dcs.fmph.uniba.sk/~obhajene/getfile.php/Ng-based-tc.pdf?

[id=1&fid=3&type=application%2Fpdf](#) - [Similar pages](#)

TCatNG Toolkit :: Text Categorization via N-Grams

TCatNG is a Java package that implement the classification technique described in Cavnar & Trenkle, "**N-Gram-Based Text Categorization**". ...

tcatng.sourceforge.net/ - 9k - [Cached](#) - [Similar pages](#)

libTextCat - Lightweight text categorization

Libtextcat is a library with functions that implement the classification technique described in Cavnar & Trenkle, "**N-Gram-Based Text Categorization**" [1]. ...

software.wise-guys.nl/libtextcat/ - 4k - [Cached](#) - [Similar pages](#)

TextCat Language Guesser

TextCat is an implementation of the **text categorization** algorithm presented in Cavnar, W. B. and J. M. Trenkle, "**N-Gram-Based Text Categorization**" In ...

www.let.rug.nl/~vannoord/TextCat/ - 4k - [Cached](#) - [Similar pages](#)

[PDF] Language Identification from Text Using N-gram Based Cumulative ...

File Format: PDF/Adobe Acrobat - [View as HTML](#)

[11] W. B. Canvar and J. M. Trenkle, **N-gram based Text Categorization**, Symposium on Document Analysis and Information Retrieval, Pages 161-176, ...

www.csis.pace.edu/~ctappert/srd2004/paper12.pdf - [Similar pages](#)

[PDF] Language and Task Independent Text Categorization with Simple ...

File Format: PDF/Adobe Acrobat - [View as HTML](#)

guage Trees and Zipping. Physical Review Letters, 88. W. Cavnar, J. Trenkle. 1994. **N-Gram-Based Text. Categorization**. Proceedings of 3rd Annual Sympo- ...

acl.ldc.upenn.edu/N/N03/N03-1025.pdf - [Similar pages](#)

Document Classification and Text Categorization

Applied Intelligence's Special Issue on **Text** and Web Mining, in press. 2001. 7. William B. Cavnar, John M. Trenkle 1994 **N-Gram-Based Text Categorization** ...

ciir.cs.umass.edu/~fuchun/readlist_all/readlist/node4.html - 12k - [Cached](#) - [Similar pages](#)

Text categorization for multi-page documents

N-Gram based text categorization. In Prof. of the 3rd Annual Symposium on Document Analysis and Information Retrieval, pages 161-175, Las Vegas, NV, 1994. ...

portal.acm.org/citation.cfm?id=379440 - [Similar pages](#)

Text classification in Asian languages without word segmentation

N-Gram-Based Text Categorization. Proceedings of SDAIR-94. 3 S. Chen and J. Goodman. 1998. An Empirical Study of Smoothing Techniques for Language Modeling. ...
portal.acm.org/citation.cfm?id=1118935.1118941 - [Similar pages](#)
[[More results from portal.acm.org](#)]

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) **[Next](#)**

Try [Google Desktop](#): search your computer as easily as you search the web.

[Search within results](#) | [Language Tools](#) | [Search Tips](#) | [Dissatisfied? Help us improve](#)

©2007 Google - [Google Home](#) - [Advertising Programs](#) - [Business Solutions](#) - [About Google](#)


[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

ngram language identification

Search

[Advanced Scholar Search](#)
[Scholar Preferences](#)
[Scholar Help](#)
Scholar [All articles](#) - [Recent articles](#) Results **1 - 10** of about **3,330** for **ngram language identification**
All Results[M Zissman](#)[J Homere](#)[E Barnard](#)[T Dunning](#)[W Cavnar](#)
Comparison of four approaches to automatic language identification of telephone speech - all 3 versions »

 MA Zissman - Speech and Audio Processing, IEEE Transactions on, 1996 -
[ieeexplore.ieee.org](#)

 ... approaches for automatic **language identification** of speech ... recognition followed by
language- dependent, interpolated **n-gram language** modeling (PRLM ...

 Cited by 118 - [Related Articles](#) - [Web Search](#) - [BL Direct](#)
Automatic language identification of telephone speech messages using phoneme recognition and N-gram ...

 MA Zissman, E Singer - Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., ..., 1994 - [ieeexplore.ieee.org](#)

... 1 0-7803-1775-0/94 \$3.00 © 1994 IEEE AUTOMATIC LANGUAGE IDENTIFICATION OF TELEPHONE

SPEECH MESSAGES USING PHONEME RECOGNITION AND N-GRAM MODELING*

Marc A. ...

 Cited by 38 - [Related Articles](#) - [Web Search](#) - [BL Direct](#)
Language identification using phoneme recognition and phonotactic language modeling - all 2 versions »

 MA Zissman - Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., ..., 1995 -
[ieeexplore.ieee.org](#)

 ... sin- gle-**language** phoneme recognilizers followed by **n-gram lan-** gilage models yielded top performance at the March 1994 NIST **language identification** evaluation. ...

 Cited by 26 - [Related Articles](#) - [Web Search](#) - [BL Direct](#)
An approach to automatic language identification based on language-dependent phone recognition - all 2 versions »

 Y Yan, E Barnard - Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., ..., 1995 -
[ieeexplore.ieee.org](#)

 ... Singer. Automatic **language identification** of telephone speech messages us- ing phoneme recognition and **N-gram** Modeling. I 1gg. International ...

 Cited by 30 - [Related Articles](#) - [Web Search](#) - [BL Direct](#)
Segment-based automatic language identification - all 13 versions »

 TJ Hazen, VW Zue - Journal of the Acoustical Society of America, 1997 - [sls.lcs.mit.edu](#)

 ... House and Neuburg showed that simple **n -gram language** models applied to error free sequences of phonetic elements as general as broad phonetic classes can ...

 Cited by 38 - [Related Articles](#) - [View as HTML](#) - [Web Search](#) - [BL Direct](#)
Automatic language identification using both N-gram and word information - all 3 versions »

BM Schulze - US Patent 6,167,369, 2000 - Google Patents

... [54] AUTOMATIC LANGUAGE IDENTIFICATION USING BOTH N-GRAM AND WORD INFORMATION ... AUTOMATIC

LANGUAGE IDENTIFICATION USING BOTH N-GRAM AND WORD INFORMATION ...

 Cited by 16 - [Related Articles](#) - [Web Search](#)

[PS] Automatic language Identification Using a Segment-Based Approach - all 12 versions »

T Hazen - 1993 - portal.acm.org

... classes is varied : : : : 42 3.8 **Language** identification accuracy of **n-gram** models using the summit phonetic recognizer with ...

Cited by 30 - [Related Articles](#) - [Web Search](#) - [Library Search](#)

[BOOK] Statistical Identification of Language - all 11 versions »

T Dunning, Computing Research Laboratory, New ... - 1994 - ling.ohio-state.edu

... **N-gram** counting using rank order statistics ... **Language** identification was done by compiling

a similar list of common short strings found in the string to be ...

Cited by 77 - [Related Articles](#) - [View as HTML](#) - [Web Search](#) - [Library Search](#)

Language identification using Gaussian mixture model tokenization - all 5 versions »

PA Torres-Carrasquillo, DA Reynolds, JR Deller Jr - Acoustics, Speech, and Signal Processing, 2002. Proceedings. ..., 2002 - ieeexplore.ieee.org

... sst.l1.mit.edu ABSTRACT Phone tokenization followed by **n-gram language** modeling has consistently provided good results for the task of **language identification**. ...

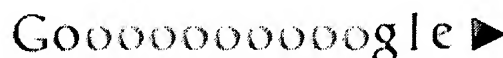
Cited by 13 - [Related Articles](#) - [Web Search](#) - [BL Direct](#)

Approaches to language identification using Gaussian mixture models and shifted delta cepstral ... - all 5 versions »

PA Torres-Carrasquillo, E Singer, MA Kohler, RJ ... - International Conference on Spoken Language Processing, 2002 - egr.msu.edu

... results indicate that automatic **language identification** (LID) systems that rely on multiple-**language** phone recognition and **n-gram language** modeling produce the ...

Cited by 19 - [Related Articles](#) - [View as HTML](#) - [Web Search](#)



Result Page: 1 2 3 4 5 6 7 8 9 10 **Next**

[Google Home](#) - [About Google](#) - [About Google Scholar](#)

©2007 Google